Session 10

STATISTICAL USES OF ADMINISTRATIVE RECORDS

Improving Data Quality Through Increased Data Sharing

Edward A. Trott[1]
Bureau of Economic Analysis

In this time of major cutbacks throughout the federal government the Bureau of Economic Analysis (BEA), along with other statistical agencies, is being asked to improve the quality of statistical information and its availability. This is a big challenge. It will require a lot of thinking "outside the lines," followed by action.

To emphasize this point, I offer the following quote;

> "In a rapidly changing world, the best solution is not to keep redesigning the organizational chart; it is to melt the rigid boundaries between organizations. The federal government should organize work according to customers' needs and anticipated outcomes, not bureaucratic turf. It should learn from America's best-run companies, in which employees no longer work in separate, isolated divisions, but in project- or product- oriented teams.
>
> To do so, the government must make three changes. It must give federal workers greater decision making authority, allowing them to operate effectively in cross-cutting ventures. It must strip federal laws of prohibitions against such cooperation. And it must order agencies to reconsider their own regulations and tradition-bound thinking."[2]

One of the recommendations of the National Performance Review is the elimination of legislative barriers to the exchange of business data among federal statistical agencies. This recommendation is referenced as DOC11 and cites the Census Bureau, Bureau of Labor Statistics, and the Bureau of Economic Analysis as participating agencies. It further recommends the reduction of reporting burden on American business.

Today I will mention several areas in BEA's international, national, and regional accounts that utilize data from the Internal Revenue Service (IRS), Bureau of the Census (BOC), and Bureau of Labor Statistics (BLS)--where we are today--and then take a look at where we could be if more barriers to information exchange are removed.

---

[1]The author is the Assistant to the Chief, Regional Economic Analysis Division, Bureau of Economic Analysis, United States Department of Commerce. He is grateful for the contributions of Ralph Kozlow, International Investment Division; Brooks Robinson and Kenneth Petrick, National Income and Wealth Division; and Wallace Bailey, Regional Economic Measurement Division to the various parts of the paper.

[2]"*Creating a Government That Works Better and Costs Less; The Report of the National Performance Review.*" Penguin Books, USA, 1993.

I.Current Uses

International Accounts

BEA and the Census are authorized to exchange certain confidential statistical data with one another for the purposes of augmenting and improving the quality of BEA's Title 22 (direct investment and international services) data and Census' Title 13 (economic census and related programs) data, under the Foreign Direct Investment and International Financial Data Improvements Act of 1990. Under the same act, the Bureau of Labor Statistics is authorized, for the purposes of augmenting and improving its data, to have access to selected information collected by BEA. Only data collected directly by BEA can be shared with Census and BLS; thus, for example, data BEA obtains from Census cannot, in turn, be shared with BLS. Similarly, only data collected directly by Census can be shared with BEA. Thus, BEA cannot have access to administrative records data obtained by Census from the IRS.

BEA and Census have made extensive use of their limited authorization to exchange data with one another and, as a result, have significantly augmented and improved the data collected in their business statistics programs. Several data exchanges have been for the purpose of obtaining detailed, establishment-level data on foreign direct investment in the United States. In 1992, BEA information on the identity of enterprises, or companies, that were foreign owned was linked to the Census' establishment-level data on the Standard Statistical Establishment List covering all U.S. establishments, to obtain detailed BOC data (State by 4-digit SIC) for 1987 on the number, employment, payroll, and shipments or sales of foreign-owned and all U.S. establishments. In 1993, BEA and Census released detailed data (mostly at the 3-digit SIC level by State) for 1989 and 1990, covering nearly all the data items on Census' Annual Survey of Manufactures (ASM) for foreign-owned and all U.S. manufacturing establishments. In 1994, BEA and Census will release detailed 1988 and 1991 ASM data for foreign-owned and all U.S. establishments. These exchanges have enabled statistical analyses that offer important new insights into the operations of foreign-owned companies and their impact on particular sectors of the U.S. economy.[3]

Other data exchanges between BEA and BOC have been for the purpose of supplementing the mailing lists for BEA's mandatory international surveys. BEA obtained confidential Census data on the identity of companies that are foreign owned, of companies that have foreign affiliates, and of companies that export services, in order to supplement mailing lists for BEA surveys of inward and outward direct investment and of international services transactions, respectively. The use of Census data to supplement the mailing lists for BEA surveys of international services transactions were particularly useful, because a significant number of potential new respondents were added

---

[3] See, for example, "Foreign Direct Investment in the United States: Establishment Data for 1987," in the October 1992 issue of the *Survey of Current Business*, and "Characteristics of Foreign-Owned U.S. Manufacturing Establishments," in the January 1994 issue of the *Survey*.

to BEA's mailing list as a result. (To further this purpose, questions on whether a given company or establishment sold, or exported, any services to foreign persons were included on a number of additional 1992 economic census surveys at BEA's request.)

In addition, BEA and the BOC are currently studying the feasibility of linking BEA's data on foreign-owned enterprises to Census' establishment-level product and product class data from the economic censuses, and product-by-country merchandise export data from its Exporter Data Base. The feasibility of other data link projects between the two Bureaus is also being discussed.

As mentioned above, in a parallel data link project, BEA's data on foreign direct investment in the United States have been linked to BLS data on all U.S. businesses. The initial results of that link, released in 1992 by BLS[4], provided data for 1989 and 1990 on the number, employment, and payroll of foreign-owned establishments. In October 1993, BLS released information on the occupational structure of foreign-owned manufacturing establishments for 1989[5].

BEA and the Department of the Treasury are also implementing limited data exchanges. BEA collects its data on direct investment and international services transactions, and Treasury collects its data on portfolio investment, under the same act (P.L. 94-473, as amended) and, in many cases, from the same U.S. companies. In order to ensure that companies are reporting correctly to both agencies, the Secretaries of Commerce and Treasury obtained OMB authorization to exchange and compare certain data. OMB approval for the data exchanges was obtained in January 1993, but BEA and Treasury are still working out the details necessary to implement the data exchanges. Note that any additional data exchanges between BEA and Treasury will require this same slow implementation process (beginning with letters from the Secretaries of Commerce and Treasury requesting OMB authorization, followed by negotiation between the agencies on the details for sharing the data.) It is hoped the time and resources required to implement data exchanges would be greatly reduced or eliminated if broad data sharing legislation were enacted.

National Accounts

BEA currently is accessing both noncorporate and corporate Statistics of Income (SOI) data via letters of agreement between IRS and BEA and the Secretary of Commerce and the IRS Commissioner as permitted by statutes. The

---

[4]United States Department of Labor. Bureau of Labor Statistics. "Employment and Wages in Foreign-Owned Businesses in the United States, 1989." Press Release, USDL 92-473 (July, 1992). and (*ibid*) "Employment and Wages in Foreign-Owned Businesses in the United States, Fourth Quarter 1990." Press Release, USDL 92-663 (October 1992).

[5]United States Department of Labor. Bureau of Labor Statistics. "New Research on Occupations in Foreign-Owned Manufacturing Establishments in the United States." Press Release, USDL 93-455 (October 1993).

Bureau's relationship with IRS has been in place for many years--dating back to 1944. At that time the Secretary of the Treasury, Henry Morganthau Jr., signed an executive order granting Commerce staff access to both noncorporate and corporate returns for statistical purposes. This executive order was superseded by statute in the early 1980's and access to SOI data from then until now has been limited in scope.

Following is a list of some of the specific topics that were recently researched via corporate return access--unless otherwise noted, adjustments to data resulting from this analysis improved the current year NIPA estimates of profits, interest, and dividends by adjusting extrapolating series to better conform to SOI definitions:

Identified discrepancies between SOI tabulations and trade association data in profits reported by life insurance companies for 1986-1987. Once identified, the SOI tabulations could be adjusted for changes in the tax law.

Identified differing reporting practices for commercial bank bad debt losses in SOI and FDIC tabulations and adjusted the current year FDIC series.

Reconciled pension and profit-sharing reporting practices between SOI and IRS Form 5500 preparatory to revising estimation methodology and data sources.

Separated business and interest receipts of credit agencies to more accurately measure interest flows.

Identified public utility joint partnership formations to ensure unduplicated reporting of new plants.

Analyzed captive finance companies' consolidation practices by parent corporations to ensure unduplicated reporting of income items.

Made adjustments to shift data for individual corporations to different industry classifications, which are more consistent with NIPA definitions and other data sources.

A secondary use of the individual company reports has been in the analysis of peculiar movements in SOI data. Because BEA has direct access to the IRS' corporate tax return information (not for the Foreign Direct Investment project), company by company panel comparisons have been compiled to review several questions regarding reporting practices, consolidation methods, or compilation errors.

Currently, actual use of noncorporate tax information has been limited to a study to determine the reporting patterns of partnership "pass-through" income reported on schedule K and by the individual partners on their Form 1040's. Results of the study led to a revision of the NIPA estimates of noncorporate partnership income and a redesign of the Form 1065 to explicitly determine the amounts of Schedule K income distributed to partners.

An additional area of data sharing between BEA and another federal agency--the Census Bureau--involves the use of the Quarterly Financial Reports (QFR) to adjust and analyze the quarterly NIPA estimates. Through a memorandum of agreement with the Census, BEA's sworn Census employees have continued to access individual company reports since 1982. Prior to 1982 BEA had access to the reports through an agreement with the Federal Trade Commission.

Primary use of this access is for the quarterly NIPA estimates. Each quarter BEA staff provide a list to Census of large non-recurring income or expense items which have been noted in a company's individual QFR. These items--mainly capital related and not associated with current production--are not treated as current income or expense items in the NIPA's. Census reviews the company report to determine if the questioned item was reported as a component of the indicator used to estimate pretax profits, or if it was excluded via several other reporting options. If the amount was included, Census identifies the exact pretax impact, which often differs from the amount in published after-tax profit reports, and the industry where the company is classified in their tabulations. The QFR also is used to augments the BEA company list with additional companies Census discovers during the review process of the quarterly reports.

In 1989 BEA entered into another cooperative arrangement with the Bureau of the Census; to explore the feasibility of using BOC data to prepare construction price indexes. In 1989, BEA staff--after being sworn in as Census agents--and the Census staff of the Construction Statistics Division (CSD) began to work on producing improved construction price indexes. The CSD provided micro-level construction data that were either collected through the Survey of Construction or the Value of New Construction Put in Place Survey. BEA staff used the data to test prospects for producing hedonic price indexes for residential and nonresidential buildings.

BEA staff produced a multifamily structures hedonic price index in 1991 for 1978-1989, and BEA began using it to prepare constant dollar estimates of the multifamily structures component of GDP at the time of the 1991 comprehensive revision of the NIPA's. Prior to this time, BEA used the BOC's Single-family Hedonic Price Index for Houses Under Construction. An article that describes the development of the index was published.[6]

Regional Accounts

For the preparation of BEA's estimates of personal income by State and county, the Regional Economic Measurement Division (REMD) receives several statistics tabulated by county from IRS Form 1040. The data are coded geographically and tabulated by the Census Bureau; the IRS, SOI Division provides these unpublished tabulations to BEA. The sample used by SOI for the Form 1040 statistics that they publish is too small to yield data for regional use. The

---

[6]See U.S. Department of Commerce, Bureau of Economic Analysis. "A Price Index for New Multifamily Housing." *Survey of Current Business*. Washington, DC: U.S. Government Printing Office, February 1993.

tabulations of dividends, taxable interest, and gross rents are the principal bases for REMD's estimates of the monetary components of property income. The tabulations of wages and salaries are vital for adjusting the components of personal income that are initially estimated on a place-of-work basis--wages and salaries, other labor income, and most of personal contributions for social insurance--to a place-of-residence basis. The tabulations of number of returns, number of exemptions, and adjusted gross income are used for REMD's quality control of the tabulations that are used directly in the estimates.

Proprietors' income and employment are important parts of the State and county estimates of personal income and employment prepared by REMD. Annual filings to the IRS by sole proprietors and partnerships are the only direct source of data for regional nonfarm proprietors' income and employment. The sample-based statistics from these filings that are prepared and published by the SOI Division are used for the calculation of national nonfarm proprietors' income. For the State and county estimates, REMD has acquired, under a contractual arrangement, individual records for sole proprietors from IRS Form 1040 Schedule C and for partnerships from IRS Form 1065. These records were then geographically and industrially coded, tabulated and combined to form the basis for REMD's State and county nonfarm proprietors' estimates. REMD-SOI data contracts have been in place for tax years 1981-83 and 1987-91.

II.Future Uses: How BEA Programs Would Be Improved Upon Implementation of DOC11

For illustrative purposes, several examples of how BEA's economic accounts can be improved with increased data sharing are given below. These are only a few of the examples that could be cited.

International Accounts

First, increased data sharing would improve the data obtained in the data link projects on foreign direct investment in the United States that BEA conducts with the BOC and BLS. As discussed earlier, BEA has shared its information on the identity of foreign-owned companies with both Census and BLS, to obtain those agencies' detailed data, by industry and State, for companies that are foreign owned. However, at present, Census and BLS cannot share their data with one another either directly or through BEA. For the data items released that were defined most similarly by both Census and BLS (number of operating establishments and employment), there are very large unexplained differences between the two agencies' data for foreign-owned establishments. For example, within manufacturing, the differences between Census and BLS data for 1990 are often at least 20 percent at the 2-digit SIC classification level; at the 3-digit and 4-digit SIC levels, the differences are often even larger. (Other industries may show even larger differences than manufacturing.) With such large unexplained differences, the results of the data link projects, and the conclusions based upon them, may be questioned. Reasons for differences may include different definitions, differences in industry coding, errors in the data, or that one agency was better able than the other agency to link its establishments to a given foreign-owned company. Because Census and BLS cannot share these data with one another currently, the exact reasons for the differences cannot now be determined. With increased data sharing, the

reasons for the differences would be learned, and appropriate corrections to the data sets, if necessary, could be made.

Second, BEA's ability to analyze the data from the BEA-Census data link project would be improved if BEA had access to IRS data on Census' file of foreign-owned establishments. BEA employees are allowed to obtain limited administrative records data directly from the IRS, but they are not allowed to access the same data at the Census Bureau, because redisclosure of administrative record data to BEA is prohibited. (For 1987, IRS data account for about 15% of the estimate of employment for all foreign-owned establishments combined.) The Commerce Department must analyze and report to Congress on the employment, market share, value added, productivity, profitability, etc. of foreign-owned business enterprises compared to all U.S. business enterprises (P.L. 101-533 section 3(c)(1)). In order that BEA may efficiently perform these analyses, they need access to the full foreign direct investment (FDI) data file at Census. (The IRS has said it will support a regulatory change that will allow BEA access to the full FDI file, but this change may take considerable time to implement and will only facilitate this one project.)

Third, full data sharing would improve BEA's direct investment data if it permitted BEA access to complete IRS data on foreign-owned U.S. businesses and on U.S. businesses that have foreign affiliates. Both BEA and the IRS have released data on the net income, assets, sales, and number of foreign-owned U.S. companies and of U.S. companies that have foreign affiliates, and the data of the two agencies frequently differ substantially. Some of these differences might be eliminated if BEA were able to obtain information from the IRS for comparing with its own information on which companies are foreign owned and which companies own foreign affiliates. Also, appropriate corrections to BEA's data sets could be made, and BEA's sample frames could be improved.

National and Regional Accounts

First, for companies that own more than one establishment, Census and BLS both annually "map" all the individual establishments on their registers to the owning enterprises. Census obtains data needed for this purpose from the mandatory annual Report of Organization survey; BLS obtains the needed data from the Multiple Worksite Report (executed by each of the States and funded by BLS) and from commercial sources, such as Dun and Bradstreet. Under data sharing, only one agency would need to gather the information and share it with the other. This change would result in one set of company/establishment relationships that BEA could use to allocate data available only for companies to an establishment industry level.

Second, REMD currently makes intensive use of tabulations of ES-202 wage and employment data in estimating the wage-related components of personal income and employment for States and counties. Any improvement in the geographic or industrial classification of the ES-202 data that might result from ES-202/Standard Statistical Establishment List integration would increase the accuracy of BEA's regional estimates.

Third, full data sharing between Census and BLS would improve the derivation of the sampling frame for BLS' Producer Price Index. BLS uses the ES-202 file for its sample frame. The price indexes would be improved if the probability of inclusion in the sample were based on its shipments or sales, because it would allow for more precise weighting of the data that are collected. Producer Price Indices are used extensively in preparing constant dollar estimates for several expenditure components of GDP and for the industry distribution of GDP.

Fourth, with full data sharing, the industry coding used for calculating both the BLS producer price indexes and the Census shipments data could be fully consistent. At present, BEA must apply BLS' price indexes, which are based upon BLS' industry codes, to Census' shipments data, which are based on Census' industry codes, in calculating real GDP by industry.

Fifth, increased data sharing will improve BEA's I-O tables and estimates of gross State product and GDP by industry. In developing its detailed industry and State estimates, BEA utilizes Census data on establishment receipts and value added, and BLS data on employment and payroll. As discussed in connection with BEA's estimates of real GDP, the accuracy and usefulness of the I-O tables and of BEA's estimates of gross State product and GDP by industry would be considerably improved through consistent industry coding of establishments. These estimates would also be improved with full BEA access to unsuppressed Census Bureau estimates of value added, cost of materials, etc.; BEA now must indirectly estimate these items for suppressed Census Bureau data cells.

Lastly, REMD has identified additional statistics from the Individual Income Tax Returns (see under "Current uses") that would strengthen its current estimates and would help to extend its current estimates to prepare for the adoption of measures more consistent with the System of National Accounts (SNA).

The transfer of these data series to BEA do not require additional legislative action, only that the data be requested from the IRS. A relatively new and growing income entry on the individual tax form--interest from municipal bonds--will yield data to replace proxy information currently used to estimate the nontaxable interest component of personal income at both the State and county level. The tabulation of pensions and annuities received would be essential for the estimation of the corresponding components of the SNA aggregates.

III Conclusions

From the current uses section of this paper, it is apparent that even the limited amount of data sharing permitted under existing legislation has already produced a remarkable increase in the amount and quality of data and analyses, with no increase in respondent burden and minimal increased costs to the statistical agencies. The examples of future uses suggest improvements in data sharing that will benefit both the customers and staff of BEA as the recommendations of DOC11 are met.

It is apparent that significant duplication and inefficiency exist in business statistics data collection programs currently. It is particularly important that we continue our efforts to promote fuller data sharing in light of the dual challenges of limited agency resources and having to monitor developments in a rapidly changing world.

The opportunity to improve the accuracy of the income and product accounts produced by BEA is here. I believe that with the implementation of DOC11, estimates prepared not only at BEA but throughout the government will be improved as to their accuracy, timeliness, and will be less costly to produce.

Health Reform Information Systems: Great Expectations, Uncertain Prospects

Edward L. Hunter
National Center for Health Statistics

Introduction

In the opening keynote address, Graham Kalton described Roger Herriot (who organized this session and died suddenly this spring) as someone that had a vision beyond what we did every day, a person that encouraged us to look up from our work occasionally and check for shifting paradigms. Roger recognized that the types of information systems proposed for health reform may well represent such a paradigm shift. The purpose of this paper is to convey a sense of the great expectations that have accompanied discussion of health reform, the importance of these proposals for research and statistics, and to discuss some of the uncertainty over whether these expectations will be realized.

Health reform

The national debate over health care reform is reaching a crescendo this summer in Washington. More than a dozen congressional committees are weighing decisions that might - or might not - lead to wholesale changes in the U.S. health care system and, as a result, in our Nation's health information systems. The work of these committees is focused on piecing together legislation that can pass both Houses before members leave the Capitol to face voters in November.

While each health reform proposal may get there by a different route, most share the same common goals: improve health insurance security for the 38 million persons without health insurance, and for those that have insurance but risk losing it due to illness or unemployment; improve access to health care; control and even reduce health care costs, and even improve the health of the American people.

Among the questions yet to be resolved:

1) Will health reform result in universal coverage for health insurance, or will more incremental approaches be adopted?

2) If there are standard health insurance packages, how generous will the benefits be?

3) How will coverage for the uninsured be financed? Options include a mandate that employers provide coverage, a mandate that individuals buy insurance, or broad-based tax increases.

4) To what extent will the Government be involved in regulating activity in the health care system?

Major proposals still under consideration include modifications to the President's proposed Health Security Act and more modest or market-oriented alternatives proposed by moderate Democrats and Republicans. At this point in the debate, it is not yet clear what type of reform - if any - will be enacted.

Role of information in the current health reform debate

It does seem clear, however, that issues of health reform will continue to dominate our agenda for the coming years regardless of what happens this session of the Congress. And it is clear that information - or the lack thereof - has become central to the emerging debate over health reform. Questions, such as who lacks health insurance and why; how and why do businesses cover employees; how is insurance related to use of services, and to health; what drives health costs up; and what will expanded benefits cost, have become central to crafting a legislative package. Yet, these are questions for which our current information systems - both survey and administrative - can provide only partial answers.

Importance of information in a reformed health system

Nearly all advocates of health reform visualize a new health care system in which information plays an increased role. Since reform will almost certainly be implemented at the State level, State governments will need better information to tailor their approaches. Health plans will need better data to select member institutions and facilities, and manage the quality and cost of patient care. Consumers will need far greater information on quality of services, satisfaction, and outcomes in order to make informed choices between competing plans. Further down the road, States will need to assure that health plans are meeting standards - and the Federal Government will need to assess the expected and unexpected impacts of reform. Finally, there will almost certainly be a debate over what went right - or wrong - that can only be conclusive if there is adequate information.

There is a growing, and encouraging, consensus among the sponsors of competing health reform proposals regarding several key information provisions. Most important among these are the need for administrative simplification - digging health providers and payers out from under the ever-growing pile of forms, and the need for greater uniformity in the way health transactions are recorded and reported. Nearly all parties to the debate recognize that we need to hold providers and health plans more accountable, that we need better information on health outcomes to guide practice, and that consumers will play a greater role as information users.

The Health Security Act: a national framework for health information

President Clinton's proposed "Health Security Act"[1] includes the most ambitious - and controversial - information proposals of any plan currently being considered. The balance of this paper will outline these proposals, and address some of the implications if current legislative deliberations result in enactment of provisions resembling these proposals.

The architects of the Health Security Act regard the availability of health information as essential to the success of managed competition, the basis for the Act. Toward that end, the Act calls for a national framework under which standardized, nationwide information would be collected for all patients, providers, and encounters with the health care system. This system would be built on a consensus around core, minimum data sets that can be used by multiple parties; standardization and simplification of currently burdensome systems of administrative and payment records; unique identification of individuals and providers to facilitate linkages; national legislation to protect the privacy of records in the system; and regional data centers to process the records and provide for access for research and statistical purposes.

Information on individuals enrolled in health insurance plans

There are three major elements in this design: enrollment, encounter, and administrative data. In the first, information would be obtained on each individual eligible for health insurance when they are enrolled in a health plan. At a minimum, this information would include identifying information (including name and address); a unique identification number; and additional demographic information such as age, sex, and (hopefully) race/ethnicity. Information on the individual's source of coverage would be included, as would any other information deemed necessary to adjust premium payments to health plans based on the individual's risk. It is possible that this list could eventually expand to include information on the individual's health status and medical history.

Data on encounters with the health care system

In the second element of the framework, a standard, minimum data set will be collected at the point of each encounter with the health care system. This data would begin with that now provided on claims forms, such as reason for visit, diagnoses, procedures, prescriptions, followup, and disposition. But since care will increasingly be delivered by prepaid systems that do not require individual claims, information would also be obtained for all such encounters. Demographic data will already have been obtained for individuals through the enrollment system, so it would not be necessary to collect such information at each encounter.

---

[1] HR 3600, "The Health Security Act of 1994", Title V.

Debate over implementing this encounter data system will focus on the length and content of the minimum data set, the extent to which an encounter might be used to obtain risk factor or other information not immediately relevant to the clinical encounter, and the types of coding and classification systems that will be used to translate recorded information into usable statistics. To make this information of maximum use for research purposes, encounter data will need to go beyond what we currently collect through the payment system, and be of better quality.

Data on plans and providers

Third, in order to facilitate the operation of the health care system, including payment of claims, certification of plans, etc., administrative data on the characteristics and operations of health plans and providers will be included in the system. This may range from the multiple affiliations that plans and providers will have with each other, to the types of services provided and patients served.

Privacy and data access

To summarize, the information framework envisioned in the Health Security Act provides for demographic information on all individuals eligible for health insurance, information on the characteristics of all health providers, and a small set of information on each encounter an individual has with the health care system.

Such a system has enormous potential - the potential to meet a wide range of research needs, and the potential - if badly designed or implemented - to jeopardize privacy protection that Americans have come to expect.

As a result, privacy protection is being given a great deal of attention by those currently debating the Health Security Act. The Administration is committed to the enactment of strong confidentiality protection,[2] and there are serious efforts being made in the Congress to craft comprehensive privacy legislation for medical records.[3]

The privacy debate over the health reform information network is critical to the eventual ability of researchers to use data that may be part of the network, for a variety of reasons. First, without addressing the public's concerns through enactment of strong protection, the Congress may not be willing to create such a network in the first place. Second, if the

---

[2] Testimony of Nan D. Hunter, Deputy General Counsel, DHHS, before the Subcommittee on Information, Justice, Transportation, and Agriculture, Committee on Government Operations, U.S. House of Representatives, April 20, 1994

[3] HR 4077, "Fair Health Information Practices Act of 1994," introduced by Representative Gary Condit (D-CA), is one example of pending legislation.

public doesn't have confidence in the privacy of their medical records, it will be impossible to assure the quality and accuracy of the information they provide to the health care system. Third, in an effort to reassure the public on privacy, it is possible that the Congress will over-react and place restrictions on access to data that will unnecessarily restrict research and statistical uses. And fourth, even with the best of intentions, it is a tough job to craft legislation that balances privacy with access to data, and we must make efforts to ensure that there are no inadvertent limits are placed on our ability to use these sources.

Statistical uses of *ADMINISTRATIVE* records?

The Federal statistical community argues that information produced through any system, including purely administrative data systems, should serve multiple purposes, and be put into a form that is conducive to research and compilation of statistics. It is an encouraging development that the Administration and the Congress are giving serious consideration to ways in which the health reform information framework can be put to maximum public use.

In fact, it is not at all clear that this system should be considered an *ADMINISTRATIVE* system - since research and statistics are among the most important uses the system is being designed to meet.

Federal statistical agencies have a strong history of making data available to the public, while protecting privacy, by producing anonymous, person level public use files. It is encouraging that this history is guiding policy development for access to data in the health reform network. As envisioned by the authors of the Health Security Act, regional data centers will be responsible for creating general-purpose files, as well as responding to individual researchers with special requirements, while being given special privacy authorities to assure the protection of individual records.

Potential uses of data from the health reform network

Public health provides a number of useful examples of how this network might serve research needs. The network will be broad and comprehensive, but will lack detailed diagnostic and clinical information typically included in patient charts. In effect, it will be a "mile wide and an inch deep." But this "mile" of data can fulfill many important public health needs, including ones we have great difficulty meeting today.

It can provide a broad picture of the population, and reasonably complete coverage of the delivery of health services. The existence of a population base for these records also provides greater opportunity to link medical events to denominators, and develop better incidence/prevalence data on certain diseases. It allows us to address subgroups of the population with greater confidence.

Examples of health applications

One specific example is the tracking of progress in improving the immunization of preschool children, a national public health priority on which we will spend hundreds of millions of dollars in the coming years. From enrollment files, we will know the ages of children that should be receiving vaccines - and will be able to use encounter data to determine immunization rates by health plans and local areas. Likewise, we will be able to monitor our success in promoting other preventive services, such as encouraging women of certain age groups to have a mammogram, and men of appropriate ages to have cholesterol screening. With this information, we will be able to better target intervention programs to prevent disease.

If a major health reform is implemented, we will also experience significant change in nearly every aspect of health care and, potentially, the health of the public. The data in this network will help us to monitor health status and outcomes, produce routine measures of quality of care, and assess changes in the organization, financing, and delivery of health services.

Potential uses in other areas

Since the network will include information on a broad population basis through the enrollment records, it may prove to be a valuable tool for demographic research. These files will, at a minimum, include a variety of demographic and geographic items, may provide us a picture of the family status of individuals, and may even record changes in individual circumstances over time.

Since enrollment for health insurance will be primarily employment-based, the network will include considerable information about the employment status and occupation of covered individuals and their family members. This may prove to have applications for labor market and employment-related research.

And, the network will include detailed information on health care providers, including facilities, professionals, and other organizations. This will provide the basis for research on the establishments and workforce of a large segment of the economy.

Potential uses for Census

The broad population coverage anticipated for the enrollment file has obvious potential uses in conducting censuses of populations. Congressional testimony by the Administration reflects current thinking that <u>some</u> access to identifiable information in the network by statistical

agencies may be appropriate.[4] As the Bureau of the Census looks toward the use of other administrative sources of data for population censuses, the health insurance enrollment file may prove to be among the most complete source of demographic information of the population. This is said, of course, with full recognition of the limitations of this system, which will be discussed briefly below.

It is possible that some provisions allowing for access to identifiable records for such purposes may be built into the health security act as it is considered by the Congress. However, it is also likely that these sorts of uses will be carefully specified to ensure that the public is not left with the perception that there will be widespread sharing of identifiable records for non-health related purposes. It is also important to remember that the Act was written with privacy protection in mind, and limits use of identifiable files to health applications except in limited circumstances.

Applications for survey research

To this point, this review has not focused on the limitations of the proposed health reform information system, but clearly, there are many. For example, it is likely that even a universal entitlement to insurance coverage will not entice all eligible individuals to participate - leaving enrollment files short of complete population coverage. We know that certain population groups are likely to be underrepresented, or excluded. We know from other administrative data sets that we have reason to suspect the quality of reporting of certain types of data.

Finally, since the framework relies principally on the reporting of encounters with the health care system, it will lack information on events, conditions, and other health issues of interest that occur outside the health care delivery system.

For the purposes of this paper, however, as we speculate on what this system might do for research and statistics, it is assumed that we can devise methods to compensate or adjust for these limitations. For example, it will be critical to augment the network with surveys and other data collection mechanisms to fill critical data gaps and provide quality checks on administrative reporting.

The information framework, then, will both rely on data collected in surveys, and can at the same time facilitate the conduct of surveys.

---

[4] Testimony of Nan D. Hunter, Deputy General Counsel, DHHS, before the Subcommittee on Census, Statistics, and Postal Personnel, Committee on Post Office and Civil Service, U.S. House of Representatives, March 16, 1994

Surveys required for health reform

At the National Center for Health Statistics, and elsewhere in the Federal statistical establishment, we also face an enormous task in developing surveys to provide information for implementing and evaluating health reform, and we have reason to hope that new systems will facilitate this work. These surveys will be essential to augment information available through the network, since no single data source will be able to provide the depth of data needed for the types of analyses we will need to be performing. We are already beginning to address some of these needs with development of new surveys, such as the National Employer Health Insurance Survey.

Sampling frames

The first, and most significant benefit of the new system will be in the construction of sampling frames. Many current surveys rely on sampling approaches that are time consuming and inefficient. The three components of the network will facilitate sampling as follows:

1) **Population sampling from the enrollment data base**, allowing for selection of sample individuals by demographic characteristics such as age, race, sex, geographic location, occupation, and possibly other proxies. Sampling and screening for rare and non-clustered population groups, currently difficult and expensive, would be greatly facilitated. Similarly, it would be possible to quickly identify and survey selected population groups, allowing for quicker turnaround for topics of current policy or research interest.

2) **Provider sampling from administrative records.** Many current health provider surveys (e.g., hospitals, physicians, nursing homes) construct sampling frames from lists provided by professional associations, phone directories, licensing agencies, or private marketing firms. These lists are subject to considerable error and possible bias, and are costly and time consuming to create. Sampling from complete and accurate lists of providers, with detailed characteristics already included in administrative files, would make provider surveys faster, less costly, and of greater quality.

3) **Sampling based on conditions, diagnoses, or procedures** that can be identified through encounter records. For example, cohorts of individuals that received specific medical treatments can be identified for followup or interviews to assess outcomes. Similarly, persons with certain diagnoses or conditions could be identified to conduct studies of access to appropriate medical care; persons with only limited use of the health care system could be identified and included in studies of barriers to access to care.

In each of these examples, it is clear that extensive research will be required to assess the extent to which sampling frames created entirely from administrative records systems will adequately represent the universe, and approaches will need to be developed to augment these frames. It is also clear that there will continue to be an important role for traditional sampling approaches where administrative frames are not adequate.

Other potential uses - record linking

Access to enrollment and encounter data in the network can also facilitate efficient linkage of detailed health status, risk, and behavior information that we will obtain through population-based surveys to the utilization and outcomes data included in encounter files.

These files can facilitate longitudinal studies, in which individuals could be enrolled in cohorts according to characteristics in the enrollment files (e.g., occupation), or based on encounters. Followup of theses cohorts could then be conducted through surveys and through analysis of encounter data.

Conclusion: prospects

To conclude, the information system described in the health security act, and touched upon in other health reform proposals, holds enormous potential for those of us interested in research and statistics. Data from enrollment, encounter, and provider data systems will augment or replace existing approaches, and provide survey researchers with new tools for sampling, conducting longitudinal followup studies, and linking survey data with outcomes. Along with a variety of new survey approaches, this information system provides us with the opportunity to fill many long-standing gaps in our understanding of the health care system and public health, and there are a variety of potential applications beyond health.

Getting there from here requires that at least four difficult questions be answered in the affirmative: 1) Can Congress actually enact a broad reform that will change the health care system? 2) In enacting any information requirements, can an appropriate balance between privacy and access to data be found that will allow the system to be useful? 3) Can all the affected interests come together to agree on standards and minimum data sets that include items of use for statistical purposes? and finally, 4) Can we develop and manage the technology for handling and protecting the volume of records that will result, and turn these records into usable statistics?

Improvements to Economic and Health Statistics: Discussion

Miron L. Straf[1]
Committee on National Statistics
National Academy of Sciences--National Research Council

It is a privilege and pleasure to discuss the two papers by Edward A. Trott, Jr., and Edward L. Hunter. Both papers are excellent and exceptionally well prepared. As such they tributes to the memory of Roger Herriot, to which we have dedicated this session. On behalf of all of us at this conference and the many others who knew or were influenced by Roger, I thank our speakers for their quality presentations.

Discussion of "Improving Data Quality Through Increased Data Sharing: The National Performance Review (NPR) Initiative," by Edward A. Trott, Jr.

Ted Trott is being modest. The data sought are even more important than one might glean from his paper. As our economy becomes an increasingly a global one, our current system of economic statistics becomes woefully out of date. Let me give you one example from the report of one of our Committee on National Statistics panels, *Behind the Numbers: U.S. Trade in the World Economy* (National Academy Press, 1992). The traditional balance of payments framework classifies transactions by geographical boundaries, so that when foreign affiliates of U.S. firms sell to other foreign firms, no effect is recorded in our trade balance (See Figure 1-1). By supplementing reports on exports and imports across geographic boundaries with information on other international business activities, the panel estimated that the difference in what the U.S. bought from and sold to foreigners in 1987 was $64 billion, less than half the reported $148 billion trade deficit for that year.

Trott shows how sharing data can improve efficiency and accuracy, but the benefits are more. In most of the cases that he mentions, more data are or would be created than the sum of what we have in separate agencies.

Sharing data also promotes new research and serves to test new theories and methods. It helps us improve our models and better understand sources of error. It furthers the use of empirical studies and other analyses in public policy formulation and evaluation. And it respects the respondents who provided the data by seeing that their information benefits society in the most effective way.

---

[1]The remarks herein are those of the author and not necessarily those of the Committee on National Statistics nor of the National Academy of Sciences--National Research Council.

Trott's paper is an excellent documentation of costly problems from not sharing and the benefits from sharing data. It also provides some innovative approaches to the problems of missing data. One example raises a question asked in the session with George Duncan and Nancy Kirkendall on confidentiality and disclosure limitation. The Census Bureau can't provide detailed micro data to the Bureau of Economic Analysis (BEA) and so suppresses some cells in the tables provided. BEA indirectly estimates these missing data. Suppose it does so through a model based on some assumptions. The Census Bureau has the original data and can determine whether the model holds or the assumptions are wrong. Should it inform the BEA?

My friends who are into Transcendental Meditation provide some important advice for missing data. Once, when I was delivering a lecture on multiple imputation methods for nonresponse, one of them intonated to the audience that the methods were unnecessary, because, if you search really hard, you can find the missing data within yourselves.

And, to me, that is Trott's message to the federal statistical system. Only, as he points out, you really don't have to search very hard.

Thus, we are led to ask if what we need is really going to be achieved by a data sharing agreement in fulfillment of recommendation 11 to the Department of Commerce from the National Performance Review. Perhaps, but not likely in our lifetime.

What is needed is a new framework to provide the incentives for finding the ways for data sharing to occur while fulfilling the intent of confidentiality law. the Committee on National Statistics explored one approach recently with a number of statistical agency heads. The basic issue is that there needs to be a quid pro quo for one party to share information with another. Truly cooperative agreements are rarely built on one-way streets. But what can be offered in return for data provided? One approach is to return analyses that depend on the data. But more can be offered.

A statistical agency can engage in designing, developing, and even managing a data collection program of another agency, so that it can obtain important data for statistical purposes. What is more the data program of the other agency could be for administrative purposes. Thus, BEA or the Census Bureau could design for the Treasury a tax data base for tax policy or even for enforcement purposes that could also provide statistical data. It's not far fetched. As David Binder has reminded us here, Statistics Canada already does it.

Discussion of "Health Reform Information Systems: Great Expectations, Uncertain Prospects," by Edward L. Hunter.

Hunter's paper is a consummate exposition of one of the greatest challenges to the federal statistical system: the information system that may arise in a reformed health care

system. Ed lays out, in as much detail as has been developed, the data to be collected and the potential uses of them. I want to complement his description with another perspective that raises some issues for the federal statistical system.

The structure of the data is in three parts: enrollment, encounter, and claims. In addition, data would be compiled on health care providers. Consumer information would be issued as some sort of report card on plans and providers. Here measures of the quality of health care services become important.

The regional data centers may be entirely new public-private ventures. They may be the places where public-use files are compiled and where linkages to other data, such as surveys, are performed. Here privacy and confidentiality are crucial.

The health care information system envisioned can provide many benefits. The information is needed for physicians to make accurate diagnoses and recommend appropriate treatment. But the information is also needed to monitor trends that affect the costs of health care, to plan for the changes in the needs for health care, and to achieve a better understanding of how behavioral and social factors are related to health and health care coverage.

It is important to note however, that, although health care information alone may provide information on health outcomes, it is not possible to understand what may have caused the outcomes unless data can be combined with other data, such as from surveys. The information serves other purposes too, including being a valuable source of social, economic, and demographic characteristics of our population. However important these purposes are, however, the health care information system would be driven by the needs to implement the health care system and not by research and statistical needs.

So that the public might benefit from access to the information for research and statistical purposes, the Committee on National Statistics expressed concerns to the Congress about legislative provisions that were proposed. As Hunter points out, some of these provisions could preclude data for important research and statistical purposes.

The Committee's first and foremost concern is that privacy and confidentiality of health care information be adequately protected. It is not necessary to sacrifice either confidentiality or the benefits of information: both are possible if legislation provides for responsible access and demonstrated, effective means to protect confidentiality. The Committee also argued that health care legislation can protect confidentiality of information and yet permit important research and statistical uses of that information by

- Prohibiting data about an individual that are collected or maintained for research and other statistical uses from being used in any administrative or enforcement action affecting that individual. This principle is referred to as *functional separation*.

- Extending confidentiality protection to identifiable data about individuals, wherever the data are maintained.

- Providing sanctions against unauthorized disclosures by any user.

- Authorizing access to health care data about individuals for research and statistical purposes whenever confidentiality can be assured.

- Creating an independent federal advisory body charged with fostering a climate of enhanced protection for all federal data about persons and responsible data dissemination for research and statistical purposes.

The Committee conveyed these points in letters to members of Congress working on health care legislation. The result is that bills were modified to permit access to health care information for public health research and for research on behavioral and social factors affecting health. Without these changes, the viability of several of our major national surveys would be threatened.

Many difficult problems remain, however. Who, for example, rules on access? One proposal is that Institutional Review Boards grant permission for access. Some parties, such as the Institute of Medicine, would not even permit access with consent. The concern is that, if access were allowed with consent, then employers might require it as a condition of employment.

Hunter talks about the system enhancing the legally mandated reporting to local and state health departments. Not everyone wants to facilitate providing this information, however, because of confidentiality concerns.

Hunter also shows the many benefits of the data in and of themselves. We must take care however, that people do not get the impression that a single, large anonymous file can serve most needs. Such a public-use tape might do so for public health purposes, but many policy purposes require different data to be combined or data to be combined in other ways. And public use tapes cannot be combined with further data that may be needed, such as information from Social Security earnings records.

The examples Hunter gives of uses for demographic research, for labor market and employment-related research, and for providing information on health care providers shows the utility of the data to other agencies: the Census Bureau, the Bureau of Labor Statistics, and the National Center for Health Statistics. What are these agencies doing to assure access to the data? I am not aware of any major initiative.

Many serious confidentiality problems remain. For example, is it appropriate to screen the system to develop a sampling frame of people with a disease like AIDS? Are

we going to allow them to be contacted? How? Should we permit contact by a call or letter that might go to a member of their household?

A specific confidentiality practice that can be harmful is when an agency specifies that information will not be used for purposes other than for which it was collected . That is a bad principle, and some bills before Congress have adopted it. Such a blanket prohibition denies many legitimate research and statistical uses. No one can foresee all potential uses of data that would benefit society. If such prohibitions were enacted and enforced, society would lack important information that it could obtain only at greater cost through new data collection that might further intrude on individual privacy.

Despite Hunter's expectations, the administration may have missed the boat on privacy and confidentiality provisions. Bills are moving now quickly through the Congress, and Representative Condit's bill can stand on its own with or without a health reform bill.

Nevertheless, the challenges for implementing such an ambitious data system are before us, and the statistical agencies can offer a special expertise here. We can look toward a health care information system to allow through research and statistics the means of providing the information required by stakeholders represented by the six P's: policy makers, public health officials, payers, providers, patients, and the public.

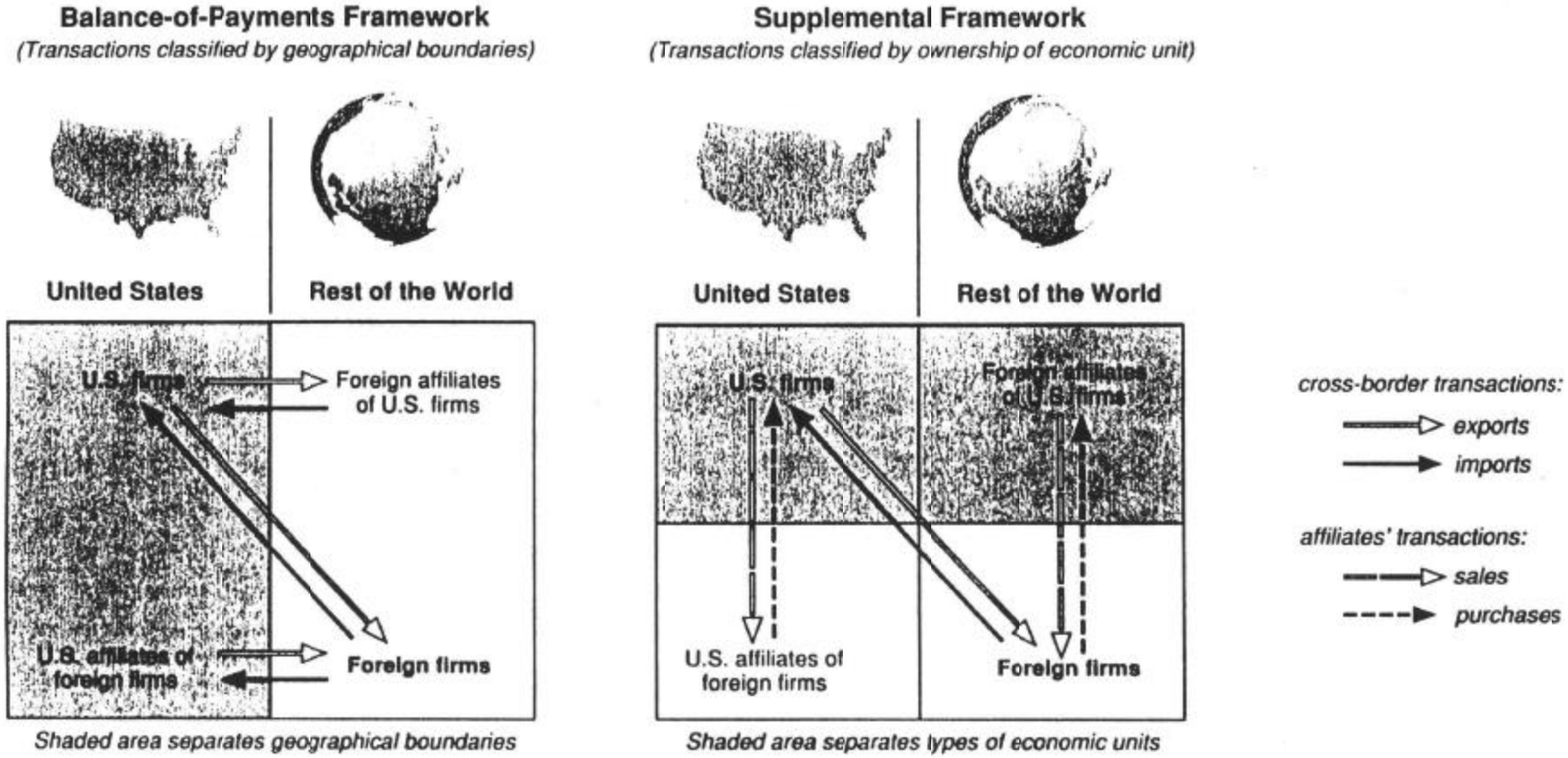

FIGURE 1-1 The balance-of-payments framework and a proposed supplemental framework.

Source: Kester, Anne Y. (1992). *Behind the Numbers: U.S. Trade in the World Economy*, Report of the Panel on Foreign Trade Statistics, Committee on National Statistics, National Academy of Sciences--National Research Council. Washington, D.C.: National Academy Press.